第27卷 第2期 工 程 数 学 学 报 Vol. 27 No. 2
2010年04月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Apr. 2010

文章编号:1005-3085(2010)02-0313-08

马氏链预测方法的统计试验研究*

夏乐天[1]， 朱元甡[2]

(1- 河海大学理学院，南京 210098； 2- 河海大学水资源环境学院，南京 210098)

摘　要: 近年来，基于绝对分布的马尔可夫链预测(ADMCP)方法、叠加马尔可夫链预测(SPMCP)方法和加权马尔可夫链预测(WMCP)方法在各种水文序列的预测理论中得到了广泛的应用。尽管各种马尔可夫链预测方法在有关应用实例中效果都不错，但是，实际上它们的预测效果是有差别的。本文通过统计试验手段对这三种马尔可夫链预测方法进行了比较研究，结论表明：WMCP方法预测精度最高，SPMCP方法预测精度次之，ADMCP方法预测精度相对来讲最差。

关键词: 马尔可夫链；水文序列；预测；统计试验

分类号: AMS(2000) 60J; 60P　　**中图分类号**: P334　　**文献标识码**: A

1 引言

近年来，张宸等[1]研究了马尔可夫链预测理论在矿区降水灾害预测中的应用；郑文瑞等[2]研究了马尔可夫链预测理论在水污染状态风险评价中的应用；冯利华等[3]用马尔可夫链预测理论研究了区域干旱的变化趋势；刘德辅等[4]用马尔可夫链预测理论研究了沿海和河口城市防灾设防标准的系统分析；宋印胜[5]和王渺林[6]研究了马尔可夫链理论在水位预测中的应用；潘俊等[7]研究了马尔可夫链理论在水库调节运行中的应用；钱家忠等[8]在地下水资源评价降水量的分析中运用了马尔可夫链预测理论；文献[9]研究了地区洪水的马尔可夫链预测；冯耀龙等[10]研究了马尔可夫链在河流丰枯状况预测中的应用；孙才志等[11]研究了马尔可夫链在降水丰枯状况预测中的应用；夏乐天等[12]研究了马尔可夫链在太湖流域和长江中下游地区的梅雨预报中的应用等。在所发表的大量研究论文中，各种马尔可夫链预测[13]方法都有不同程度的应用，但是常用的三种马尔可夫链预测方法的优劣性却一直鲜有讨论。

由于我国有关水文计算的规范规定：我国主要河流的径流量和降水量均可假定为服从P-III型分布或三参数对数正态分布。因此本文的主要目的是利用随机模拟技术，生成服从P-III型分布和三参数对数正态分布的相依伪随机数序列，然后再对各种不同的马尔可夫链预测方法进行比较分析并讨论各种因素对预测精度的影响。所讨论的主要问题是：

1) 各种马尔可夫链预测方法的适应性，包括对不同总体分布的适应性(变量总体分布为对数正态与P-III型)，和对不同的序列生成模型的适应性；

2) 讨论各种马尔可夫链预测方法的有效性，得出哪种预测方法较优的结论。

2 试验设计

2.1 伪随机数序列的长度和分布参数的设计

收稿日期: 2008-09-03. **作者简介**: 夏乐天(1956年1月生)，男，博士，教授. 研究方向：应用概率统计与随机水文学.
***基金项目**: 河海大学自然科学基金 (2084408319).

由于天然河流的实测水文资料比较短，一般少于60年，因此只有在小样本时讨论预测方法的优劣，才会有实际的水文意义。出于这种考虑，本文对于生成的序列，仅考虑长度为100的服从P-III型分布和三参数对数正态分布的相依伪随机数序列，利用适当的分组方法确定这些数据的状态。假定实测资料的长度为60年，利用这60年的资料所对应的状态预测出未来40年的状态，然后与这40年的理论状态做比较，为消除偶然性，可由生成的序列，连续考虑长度为100的服从P-III型分布和三参数对数正态分布的相依伪随机数序列，分别利用前60年的资料所对应的状态预测出未来40年的状态，然后与后40年的理论状态做比较，算出平均误差，最后定出各种马尔可夫链预测方法的优劣。为在一定的参数范围内讨论评价各种马尔可夫链预测方法的优劣，共采用了5组总体参数，见表1。

表1: 总体参数值表

序号	EX	C_v	C_s	$\rho_x(1)$
1	1000	0.25	1.00	0.2
2	1000	0.50	1.50	0.2
3	1000	0.50	2.00	0.2
4	1000	1.00	2.50	0.2
5	1000	1.00	3.50	0.2

经验表明：大多数实际径流量和降水量资料的参数是在表1的参数值的范围内，序列的一阶相关系数统一规定为0.2，由于一般的水文序列均为弱相关序列。

2.2 平稳马尔可夫链的模拟

许多水文过程都表现出明显的序列相关，通常以$\rho_x(k)$表示随机变量X的k阶序列相关系数，如果可以把$\rho_x(k)$近似地表示成$(\rho_x(1))^k$，则随机变量X的时间序列可以用一阶平稳马尔可夫链作为模型[14]。

1961年，在径流调节的研究中开始使用一阶Markov模型，即考虑序列相邻项的相关关系以生成水文序列[15]。直到目前，这种模型仍是水文学中使用最多的一种，故本节对该种模型生成的序列给出较为详尽的介绍。

确切地说，一阶马尔可夫链由下式定义

$$x_{i+1} = \mu_x + \rho_x(1)(x_i - \mu_x) + \varepsilon_{i+1}, \tag{1}$$

其中x_i是马尔可夫链在时刻i处的值，μ_x是X的均值，$\rho_x(1)$是一阶序列的自相关系数，$\{\varepsilon_i\}$是满足$E\varepsilon_i = 0$, $\mathrm{Var}\varepsilon_i = \sigma_\varepsilon^2$的独立随机变量序列。这一模型表明：$X$在某一时刻的取值，只与其前一时刻的取值有关，再加上一个随机变量。设以σ_x^2表示X的方差，考虑到ε_{i+1}与X_i相互独立，故可得σ_x^2与σ_ε^2的关系如下

$$\sigma_x^2 = \mathrm{Var}(X_{i+1}) = E[X_{i+1} - \mu_x]^2 = \rho_x^2(1)E(X_i - \mu_x)^2 + E\varepsilon_{i+1}^2 = \rho_x^2(1)\sigma_x^2 + \sigma_\varepsilon^2,$$

即

$$\sigma_\varepsilon^2 = \sigma_x^2\left[\,1 - \rho_x^2(1)\,\right]. \tag{2}$$

假设X的分布是正态的$N(\mu_x, \sigma_x^2)$，则由(1)可知ε的分布是$N(0, \sigma_\varepsilon^2)$。设以$r$表示标准正态随机变量，则$r\sigma_x\sqrt{1-\rho_x^2(1)}$是服从分布$N(0, \sigma_\varepsilon^2)$的随机变量。于是要抽样服从分

布 $N(\mu_x,\sigma_x^2)$、一阶序列的自相关系数是 $\rho_x(1)$、且符合一阶Markov模型的序列 $\{X_i\}$，其抽样公式应如下式所示

$$x_{i+1}=\mu_x+\rho_x(1)(x_i-\mu_x)+r_{i+1}\sigma_x\sqrt{1-\rho_x^2(1)}. \tag{3}$$

为利用上式生成 x_{i+1}，首先要使用某种方法得出 μ_x, σ_x, $\rho_x(1)$ 的估计值 $\bar{x}$, s_x, $r_x(1)$。$\bar{x}$, s_x 的估计值皆为已知，$r_x(1)$ 可按自相关系数的定义采用下式计算

$$\begin{aligned}r_x(1)&=\frac{\frac{1}{n-1}\sum\limits_{i=1}^{n-1}x_ix_{i+1}-\frac{1}{(n-1)^2}\sum\limits_{i=1}^{n-1}x_i\sum\limits_{i=1}^{n-1}x_{i+1}}{\left[\frac{1}{n-1}\sum\limits_{i=1}^{n-1}x_i^2-\frac{1}{(n-1)^2}\Big(\sum\limits_{i=1}^{n-1}x_i\Big)^2\right]^{1/2}\left[\frac{1}{n-1}\sum\limits_{i=1}^{n-1}x_{i+1}^2-\frac{1}{(n-1)^2}\Big(\sum\limits_{i=1}^{n-1}x_{i+1}\Big)^2\right]^{1/2}}\\&=\frac{\sum\limits_{i=1}^{n-1}x_ix_{i+1}-\frac{1}{n-1}\sum\limits_{i=1}^{n-1}x_i\sum\limits_{i=1}^{n-1}x_{i+1}}{\left[\sum\limits_{i=1}^{n-1}x_i^2-\frac{1}{n-1}\Big(\sum\limits_{i=1}^{n-1}x_i\Big)^2\right]^{1/2}\left[\sum\limits_{i=1}^{n-1}x_{i+1}^2-\frac{1}{n-1}\Big(\sum\limits_{i=1}^{n-1}x_{i+1}\Big)^2\right]^{1/2}},\end{aligned} \tag{4}$$

X_i 的初值 x_1，可以由分布 $N(\mu_x,\sigma_x^2)$ 抽得。为了消除初值的影响，可以将生成序列的前几项 (例如前50项至100项) 去掉。

式 (3) 曾被广泛地用于生成年径流序列。但由于正态分布可取负值的特性与径流过程的特性 (不可能出现负值) 有差异，特别是 μ_x 较小而 σ_x 较大时，用 (3) 式生成的 x_i 常会出现负值。如负值出现的机会不多，则可只用这负的 x_i 来生成下一个 x_{i+1}，然后将 x_i 从序列中去掉舍弃不用。但当负的 x_i 出现很多时，则表明径流过程 X_i 的分布，与正态分布差异显著，此时必须考虑其它分布。在水文统计中常用的分布有两种，即Gamma分布或P-III型分布及对数正态分布。

为使生成的 X 能服从P-III型分布时，可使用下式[14]

$$x_{i+1}=\mu_x+\rho_x(1)(x_i-\mu_x)+\xi_{i+1}\sigma_x\sqrt{1-\rho_x^2(1)}, \tag{5}$$

其中

$$x_{i+1}=\frac{2}{C_{s_\xi}}\Big(1+\frac{C_{s_\xi}r_{i+1}}{6}-\frac{C_{s_\xi}^2}{36}\Big)^3-\frac{2}{C_{s_\xi}}, \tag{6}$$

r_{i+1} 为 $N(0,1)$ 随机数，C_{s_ξ} 为 ξ 的偏态系数，由下式给出

$$C_{s_\xi}=\frac{1-\rho_x^2(1)}{[1-\rho_x^2(1)]^{3/2}}C_{s_x}. \tag{7}$$

为使生成的 X 能服从三参数对数正态分布，可假定 $Y=\ln(X-a)$ 服从正态分布 $N(\mu_y,\sigma_y^2)$，于是 Y 可采用 (3) 式的生成方法，即

$$y_{i+1}=\mu_y+\rho_y(1)(y_i-\mu_y)+r_{i+1}\sigma_y\sqrt{1-\rho_y^2(1)}, \tag{8}$$

其中 μ_y, σ_y, $\rho_y(1)$ 分别为 Y 的均值、均方差和一阶序列的自相关系数。直接用 x 表示，这一生成方法为

$$x_{i+1}=a+\big\{\exp[\mu_y(1-\rho_y(1))]\big\}(x_i-a)^{\rho_y(1)}\delta_{i+1}, \tag{9}$$

其中

$$\delta_{i+1}=\exp\big\{[1-\rho_y^2(1)]^{1/2}\sigma_yr_{i+1}\big\}.$$

具体计算时，首先用某种方法估计出 X 的有关参数 $\mu_x, \sigma_x, C_{s_x}, \rho_x(1)$ (其中 C_{s_x} 为 X 的偏态系数)，再由 X 的前二阶矩与参数间的关系[14]，可得如下几个公式

$$\sigma_y = \sqrt{\ln(1-\eta^2)}, \quad a = \mu_x(1 - C_{v_x}/\eta),$$

$$\eta = \Big[\frac{\sqrt{C_{s_x}^2+4}+C_{s_x}}{2}\Big]^{1/3} - \Big[\frac{\sqrt{C_{s_x}^2+4}-C_{s_x}}{2}\Big]^{1/3},$$

$$\mu_y = \ln(\mu_x - a) - \frac{1}{2}\ln(1+\eta^2), \quad \rho_y(1) = \frac{1}{\sigma_y^2}\ln[(\mathrm{e}^{\sigma_y^2}-1)\rho_x(1)+1].$$

将这些算得的参数代入(8)式或(9)式，则所生成的序列 $\{X_i\}$，将能保持 X 的给定参数：μ_x, σ_x, C_{s_x} 以及 $\rho_x(1)$。

3 计算方案的设计与统计试验成果的综合分析

3.1 计算方案、预测偏差分析与统计试验成果

3.1.1 计算方案与预测偏差

根据(5)式和(9)式，按表1中给定的各组参数分别生成长度为100000的服从P-III型分布和对数正态分布的马尔可夫链序列 x_i，依次取100个数据(分别考虑取50次、200次和1000次)，用三种马尔可夫链预测方法分别去做预测。即假定实测资料的长度为60年，用文献[13]3.3.1中介绍的分级方法，对一组给定的参数 $\alpha_1, \alpha_2, \alpha_3$ 和 α_4，把这60年的指标值分级，利用这60年的资料所对应的状态分别用三种马尔可夫链预测方法预测出未来40年的状态(共做50次、200次和1000次)；然后把逐步预测出的状态与理论状态作差，对这个差的绝对值先按次数作均值和均方差，代表逐步预测偏差和逐步预测偏差均方差，再考虑逐步预测偏差和逐步预测偏差均方差的平均值，可称为预测偏差均值和预测偏差均方差，这个预测偏差均值和预测偏差均方差就是度量预测方法优劣的两个标准。显然，预测偏差均值和预测偏差均方差的值是越小越好，它们反映了预测的精度。这个过程可用数学语言表述如下。

对于100个生成的服从某种分布(P-III型分布或对数正态分布)的数据 $x_1,\cdots,x_{60},\cdots,x_{100}$，把后面40个数据对应的状态记为 $z_1,\cdots,z_{40}$，作为理论数据，利用前面60个数据 $x_1,\cdots,x_{60}$ 所对应的状态，通过某种马尔可夫链预测方法逐步预测出后面40个数据所对应的状态记为 $z_1^{(j)},\cdots,z_{40}^{(j)}$，$j=1,\cdots,c$，其中 c 为模拟次数，则第 i 步逐步预测偏差定义为

$$m\overline{z}_i = \frac{1}{c}\sum_{j=1}^{c}\left| z_i^{(j)} - z_i \right|, \quad i=1,\cdots,40, \tag{10}$$

预测偏差均值定义为

$$m\overline{z} = \frac{1}{40}\sum_{i=1}^{40} m\overline{z}_i = \frac{1}{40c}\sum_{i=1}^{40}\sum_{j=1}^{c}\left| z_i^{(j)} - z_i \right|, \tag{11}$$

第 i 步逐步预测偏差均方差定义为

$$s\overline{z}_i = \sqrt{\frac{1}{c-1}\sum_{j=1}^{c}\Big(\left|z_i^{(j)} - z_i\right| - m\overline{z}_i\Big)^2}, \quad i=1,\cdots,40, \tag{12}$$

预测偏差均方差定义为

$$s\overline{z} = \frac{1}{40}\sum_{i=1}^{40} s\overline{z}_i\ . \tag{13}$$

3.1.2 预测成果汇编

取定 $EX = 1000$, $\rho_x(1) = 0.2$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1.2, 0.6, 0.4, 1.0)$，把两种分布的预测统计试验简要结果整理如下，见表2和表3。

表2: P-III型分布预测统计试验简要结果

模拟次数	总体参数		ADMCP 法		SPMCP 法		WMCP 法	
c	C_v	C_s	$m\overline{z}$	$s\overline{z}$	$m\overline{z}$	$s\overline{z}$	$m\overline{z}$	$s\overline{z}$
50	0.25	1.00	0.9840	0.8549	0.9201	0.8365	0.9082	0.8210
	0.50	1.50	0.9493	0.8518	0.9000	0.8275	0.8831	0.8204
	0.50	2.00	0.8088	0.7988	0.7864	0.7848	0.7879	0.7775
	1.00	2.50	0.7472	0.7841	0.7233	0.7617	0.7175	0.7650
	1.00	3.50	0.7521	0.7927	0.7083	0.7895	0.6986	0.7951
	平均值		0.8483	0.8165	0.8076	0.8000	0.7991	0.7958
	精度提高(%)		0.00	0.00	4.79	2.02	5.80	2.53
200	0.25	1.00	0.9487	0.8372	0.8975	0.8255	0.8766	0.8085
	0.50	1.50	0.8946	0.8182	0.8504	0.8037	0.8378	0.7950
	0.50	2.00	0.8208	0.8092	0.7823	0.7873	0.7785	0.7830
	1.00	2.50	0.8044	0.8069	0.7836	0.7947	0.7743	0.7960
	1.00	3.50	0.7322	0.7945	0.6849	0.7672	0.6898	0.7616
	平均值		0.8401	0.8132	0.7997	0.7957	0.7914	0.7888
	精度提高(%)		0.00	0.00	4.81	2.15	5.80	3.00
1000	0.25	1.00	0.9556	0.8488	0.8986	0.8277	0.8782	0.8129
	0.50	1.50	0.9007	0.8190	0.8573	0.8095	0.8439	0.7965
	0.50	2.00	0.8534	0.8188	0.8191	0.8068	0.8075	0.7992
	1.00	2.50	0.7875	0.7980	0.7500	0.7870	0.7426	0.7789
	1.00	3.50	0.6996	0.7826	0.6710	0.7737	0.6691	0.7677
	平均值		0.8394	0.8134	0.7992	0.8009	0.7883	0.7910
	精度提高(%)		0.00	0.00	4.78	1.54	6.09	2.75

3.2 统计试验成果分析

3.2.1 三种马尔可夫链预测方法的预测精度不同

就参数 (C_v, C_s) 的三组值的平均情况来说，三种马尔可夫链预测方法的预测精度是不同的(见表2、表3中各部分的最后二行)，具体地说：加权马尔可夫链预测方法的预测精度最高；叠加马尔可夫链预测方法的预测精度次之；基于绝对分布的马尔可夫链预测方法的预测精度最低。

事实上，对P-III型分布而言，以基于绝对分布的马尔可夫链预测方法(ADMCP法)为基准，叠加马尔可夫链预测方法(SPMCP法)的预测偏差均值要提高 $4.78 \sim 4.81$ 个百分点，预测偏差均方差要提高 $1.54 \sim 2.15$ 个百分点，加权马尔可夫链预测方法(WMCP法)的预测偏差均值要提高 $5.80 \sim 6.09$ 个百分点，预测偏差均方差要提高 $2.53 \sim 3.00$ 个百分点。

对三参数对数正态分布而言，以基于绝对分布的马尔可夫链预测方法(ADMCP法)为基准，重复试验200次以上，叠加马尔可夫链预测方法(SPMCP法)的预测偏差均值要提高4.69 ~ 5.54个百分点，预测偏差均方差要提高1.71 ~ 2.19个百分点；加权马尔可夫链预测方法(WMCP法)的预测偏差均值要提高5.82 ~ 6.92个百分点，预测偏差均方差要提高3.14 ~ 3.51个百分点。

表3: 对数正态分布预测统计试验简要结果

模拟次数	总体参数		ADMCP法		SPMCP法		WMCP法	
c	C_v	C_s	$m\bar{z}$	$s\bar{z}$	$m\bar{z}$	$s\bar{z}$	$m\bar{z}$	$s\bar{z}$
50	0.25	1.00	1.0046	0.8701	0.9434	0.8408	0.9372	0.8336
	0.50	1.50	0.9290	0.8364	0.8842	0.8156	0.8696	0.7976
	0.50	2.00	0.8774	0.8166	0.8415	0.7835	0.8217	0.7763
	1.00	2.50	0.7828	0.7940	0.7650	0.7791	0.7517	0.7705
	1.00	3.50	0.7917	0.7735	0.7458	0.8017	0.7500	0.7841
	平均值		0.8483	0.8771	0.8181	0.8360	0.8041	0.8260
	精度提高(%)		0.00	0.00	0.00	4.69	1.71	5.82
200	0.25	1.00	0.9315	0.8379	0.8693	0.8135	0.8568	0.8042
	0.50	1.50	0.9013	0.8270	0.8578	0.8081	0.8406	0.7937
	0.50	2.00	0.8608	0.8197	0.7936	0.7946	0.7872	0.7837
	1.00	2.50	0.8234	0.8085	0.7929	0.7991	0.7844	0.7901
	1.00	3.50	0.7460	0.7878	0.7135	0.7763	0.6992	0.7658
	平均值		0.8526	0.8162	0.8054	0.7983	0.7936	0.7875
	精度提高(%)		0.00	0.00	5.54	2.19	6.92	3.15
1000	0.25	1.00	0.9501	0.8461	0.8911	0.8247	0.8706	0.8096
	0.50	1.50	0.9007	0.8190	0.8573	0.8095	0.8439	0.7965
	0.50	2.00	0.8390	0.8106	0.7983	0.7870	0.7813	0.7751
	1.00	2.50	0.7967	0.7977	0.7542	0.7796	0.7449	0.7696
	1.00	3.50	0.7221	0.7780	0.6880	0.7622	0.6796	0.7537
	平均值		0.8771	0.8181	0.8360	0.8041	0.8260	0.7924
	精度提高(%)		0.00	0.00	4.69	1.71	5.82	3.14

3.2.2 统计试验的次数c对预测结果有一定的影响

事实上试验的次数过少，例如$c = 50$时，预测精度较低，$c \geq 200$时，预测精度趋向稳定。

3.2.3 各种马尔可夫链预测方法的预测精度具有对不同分布的适应性

以试验次数 $c = 200$为例， 基于P-III型分布的五组参数的平均预测偏差均值为ADMCP法0.8401，SPMCP法0.7997，WMCP法0.7914，基于对数正态分布的五组参数的平均预测偏差均值为ADMCP法0.8526，SPMCP法0.8054，WMCP法0.7936。

基于P-III型分布的五组参数的平均预测偏差均方差为ADMCP法0.8132，SPMCP法0.7957，WMCP法 0.7888，基于对数正态分布的五组参数的平均预测偏差均方差为 ADMCP法0.8162，SPMCP法0.7983，WMCP法0.7875。可见基于两种分布的预测精度没有本质的差别。

3.2.4 参数 (C_v, C_s) 的不同对预测精度的影响显著

一般来说，随着 C_v 和 C_s 的增大，预测偏差呈下降趋势，且 WMCP 法的预测结果最为稳健。

以试验次数 $c = 200$ 为例，基于 P-III 型分布的五组参数，ADMCP 法的 $\max m\overline{z} = 0.9487$，$\min m\overline{z} = 0.7322$，$\max s\overline{z} = 0.8372$，$\min s\overline{z} = 0.7945$；SPMCP 法的 $\max m\overline{z} = 0.8975$，$\min m\overline{z} = 0.6849$，$\max s\overline{z} = 0.8255$，$\min s\overline{z} = 0.7672$；WMCP 法的 $\max m\overline{z} = 0.8766$，$\min m\overline{z} = 0.6898$，$\max s\overline{z} = 0.8085$，$\min s\overline{z} = 0.7616$。$m\overline{z}$ 的极差在 $0.1868 \sim 0.2165$，$s\overline{z}$ 的极差在 $0.0469 \sim 0.0583$，一般说来，ADMCP 法的极差大一些，SPMCP 法的极差略小，WMCP 法的的极差最小，这说明：在参数 (C_v, C_s) 变化时，WMCP 法的预测结果最为稳健。

基于对数正态分布的五组参数，ADMCP 法的 $\max m\overline{z} = 0.9315$，$\min m\overline{z} = 0.7462$，$\max s\overline{z} = 0.8379$，$\min s\overline{z} = 0.7878$；SPMCP 法的 $\max m\overline{z} = 0.8693$，$\min m\overline{z} = 0.7135$，$\max s\overline{z} = 0.8135$，$\min s\overline{z} = 0.7763$；WMCP 法的 $\max m\overline{z} = 0.8568$，$\min m\overline{z} = 0.6992$，$\max s\overline{z} = 0.8042$，$\min s\overline{z} = 0.7658$。$m\overline{z}$ 的极差在 $0.1558 \sim 0.1863$，$s\overline{z}$ 的极差在 $0.0384 \sim 0.0501$，一般说来也是，ADMCP 法的极差大一些，SPMCP 法的极差略小，WMCP 法的极差最小，这说明：在参数 (C_v, C_s) 变化时，WMCP 法的预测结果也最为稳健。

由于篇幅所限，我们将另外撰文进一步讨论 C_v 固定，C_s 逐步增加或 C_s 固定，C_v 逐步增加对预测精度的影响，以及样本容量、一步自相关系数 $\rho_x(1)$、指标值分级的不同对预测精度的影响。

4 小结

本文通过统计试验对各种不同的马尔可夫链预测方法进行了比较分析并讨论了各种因素对预测精度的影响。

统计试验结果表明：

1) 加权马尔可夫链预测方法的预测精度最高；叠加马尔可夫链预测方法的预测精度次之；基于绝对分布的马尔可夫链预测方法的预测精度最低。

2) 统计试验的次数 c 对预测结果有一定的影响，当 $c \geq 200$ 时，预测精度趋向稳定。

3) 各种马尔可夫链预测方法的预测精度具有对不同分布的适应性。

进一步讨论可参见文献[13]。

参考文献：

[1] 张宸，林启太. 模糊马尔可夫链状模型在矿区降水灾害预测中的应用[J]. 国外建材科技，2004, (1): 56-58
Zhang C, Lin Q T. Application of fuzzy Markov chain model in predicting the precipitation state of mineral region[J]. Science and Technology of Overseas Building Materials, 2004, (1): 56-58

[2] 郑文瑞，王新代，纪昆等. 非确定数学方法在水污染状态风险评价中的应用[J]. 吉林大学学报，2003, (1): 59-62
Zheng W R, Wang X D, Ji K, *et al*. Application of undetermined mathematical method in risk assessment of water pollution conditions[J]. Journal of Changchun University of Science and Technology, 2003, (1): 59-62

[3] 冯利华，陈雄. 浙江干旱的变化趋势[J]. 农业系统科学与综合研究，2001, (3): 177-179
Feng L H, Chen X. Change tendency of drought in Zhejiang province[J]. System Sciences and Comprehensive Studies in Agriculture, 2001, (3): 177-179

[4] 刘德辅，褚晓明，王树青. 沿海和河口城市防灾设防标准系统分析[J]. 灾害学，2001, (4): 1-7
Liu D F, Chu X M, Wang S Q. Systematic analysis on standards of disaster fortification for cities along coast and river mouth[J]. Journal of Catastrophology, 2001, (4): 1-7

[5] 宋印胜. 马尔可夫链模型在地下水水位预测中的应用[J]. 山东地质，1998, (1): 34-40
Song Y S. Application of Markov-chain on the prediction of underground water levels[J]. Geology of Shandong, 1998, (1): 34-40
[6] 王渺林. 灰色马尔可夫模型在寸滩站年最高水位预测中的应用[J]. 四川水利，2004, (2): 10-11
Wang M L. Application of fuzzy Markov chain model on the prediction of the highest water levels at Cuntan Station[J]. Hydraulic Engineering of Sichuan, 2004, (2): 10-11
[7] 潘俊，陈欣，孙才志. 年调节平原水库与地下水联合调蓄分析[J]. 水文地质工程地质，2004, (2): 67-71
Pan J, Chen X, Sun C Z. Research on the conjunctive regulation of an annual regulating plain reservoir and groundwater[J]. Hydrogeology and Engineering Geology, 2004, (2): 67-71
[8] 钱家忠，朱学愚，吴剑锋. 地下水资源评价中降水量的时间序列-马尔可夫模型[J]. 地理科学，2001, (4): 350-353
Qian J Z, Zhu X Y, Wu J F. Time series-Markov prediction model for precipitation in the course of evaluation of groundwater resources[J]. Scientia Geographica Sinica, 2001, (4): 350-353
[9] Sen Z K. Critical drought analysis by second order Markov chain[J]. Journal of Hydrology, 1990, 120(1-4): 183-202
[10] 冯耀龙，韩文秀. 权马尔可夫链在河流丰枯状况预测中的应用[J]. 系统工程理论与实践，1999, (10): 89-93
Feng Y L, Han W X. The application of weighted Markov-chain to the prediction of river runoff state[J]. Journal of Systems Engineering Theory & its Application, 1999, (10): 89-93
[11] 孙才志，张戈，林学钰. 加权马尔可夫链在降水丰枯状况预测中的应用[J]. 系统工程理论与实践，2003, (4): 100-105
Sun C Z, Zhang G, Lin X Y. Model of Markov chain with weights and its application in predicting the precipitation state[J]. Journal of Systems Engineering Theory & its Application, 2003, (4): 100-105
[12] 夏乐天. 梅雨强度的指数权马尔可夫链预测[J]. 水利学报，2005, (8): 988-993
Xia L T. Prediction of plum rain intensity based on index weighted Markov chain[J]. Journal of Hydraulic Engineering, 2005, (8): 988-993
[13] 夏乐天. 马尔可夫链预测方法及其在水文序列中的应用研究[D]. 河海大学博士学位论文，2005: 60-66
Xia L T. Research of Markov chain prediction method and its application on hydrology series[D]. Phd dissertation of Hohai University, 2005: 60-66
[14] 丛树铮等. 水文学中的概率统计基础[M]. 北京：水利出版社，1981: 378-384
Cong S Z, *et al.* Probability Therory and Statistics in Hydrology[M]. Beijing: China Hydrology Press, 1981: 378-384
[15] Fiering M B. Queuing theory and simulation in reservoir design[J]. Proceedings American Society of Civil Engineers, 1961, 87(HY6): 39-68

Study on Statistical Experiment of Markov Chain Prediction Methods

XIA Le-tian[1], ZHU Yuan-shen[2]

(1- College of Science, Hohai University, Nanjing 210098;
2- College of Hydrology and Water Resources, Hohai University, Nanjing 210098)

Abstract: In recent years, the weighted Markov chain prediction method (WMCP), the Markov chain prediction method based on absolute distribution (ADMCP) and probability summation (SPMCP) have been widely used in the various hydrology series. Although the application results of various Markov chain prediction methods in real situations are seemed good, the prediction accuracy of these three kinds of Markov chain prediction methods is different. The three prediction methods are studied by using statistical experiment. The experiment results indicate that the prediction accuracy of the WMCP method is the highest, the ADMCP method is the lowest, and the SPMCP method is in the middle.

Keywords: Markov chain; hydrology series; prediction; statistical experiment

Received: 03 Sep 2008. **Accepted:** 31 Dec 2009.
Foundation item: The Natural Science Foundation of Hohai University (2084408319).